

Alignement avec brèches

Thierry Lecroq
Université de Rouen

Coût affine sur les brèches

Une brèche de longueur k est pénalisée par :

$$br(k) = g + h \times (k-1)$$

- g : coût d'ouverture
- h : coût d'extension

Programmation dynamique

On utilise 3 tables à $m+1$ lignes et $n+1$ colonnes D , I et T :

$$D[i,j] = \min \{D[i-1,j]+h, T[i-1,j]+g\}$$

$$I[i,j] = \min \{D[i,j-1]+h, T[i,j-1]+g\}$$

$$T[i,j] = \min \{T[i-1,j-1]+Sub(x[i],y[j]), D[i,j], I[i,j]\}$$

pour $0 \leq i \leq m-1$ et $0 \leq j \leq n-1$.

Initialisations marginales

$$D[-1,-1] = D[i,-1] = D[-1,j] = \infty$$

$$I[-1,-1] = I[i,-1] = I[-1,j] = \infty$$

pour $0 \leq i \leq m-1$ et $0 \leq j \leq n-1$ et

$$T[-1,-1] = 0, T[0,-1] = T[-1,0] = g$$

$$T[i,-1] = T[i-1,-1] + h$$

$$T[-1,j] = T[-1,j-1] + h$$

pour $1 \leq i \leq m-1$ et $1 \leq j \leq n-1$.

- $D[i,j]$: coût d'un meilleur alignement entre $x[0..i]$ et $y[0..j]$ se terminant par des suppressions de lettres de x ;
- $I[i,j]$: coût d'un meilleur alignement entre $x[0..i]$ et $y[0..j]$ se terminant par des insertions de lettres de y ;
- $T[i,j]$: coût d'un meilleur alignement entre $x[0..i]$ et $y[0..j]$.

algo BRECHE(x, m, y, n)

pour $i \leftarrow -1$ **à** $m-1$ **faire**

$(D[i,-1], I[i,-1]) \leftarrow (\infty, \infty)$

$T[-1,-1] \leftarrow 0$

$T[0,-1] \leftarrow g$

pour $i \leftarrow -1$ **à** $m-1$ **faire**

$T[i,-1] \leftarrow T[i-1,-1] + h$

$T[-1,0] \leftarrow g$

pour $j \leftarrow 1$ **à** $n-1$ **faire**

$T[-1,j] \leftarrow T[-1,j-1] + h$

pour $j \leftarrow 0$ **à** $n-1$ **faire**

$(D[-1,j], I[-1,j]) \leftarrow (\infty, \infty)$

pour $i \leftarrow 0$ **à** $m-1$ **faire**

$D[i,j] \leftarrow \min \{D[i-1,j]+h, T[i-1,j]+g\}$

$I[i,j] \leftarrow \min \{D[i,j-1]+h, T[i,j-1]+g\}$

$T[i,j] \leftarrow \min \{T[i-1,j-1]+Sub(x[i],y[j]), D[i,j], I[i,j]\}$

retourner $T[m-1,n-1]$

- complexité : $O(mn)$ en temps, $O(\min\{m,n\})$ en espace si seul $T[m-1,n-1]$ doit être retourner ;
- tracé arrière délicat.

Référence

O. Gotoh,

An improved algorithm for matching biological sequences,

Journal of Molecular Biology, **162**:705-708,
1982.